

5

## RELATIVE RANGE CAMERA CALIBRATION

by

10 John C. Krumm

### BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates in general to range imaging systems and more particularly to a method and a system for measuring a relative position and orientation of range cameras using a movement of an object within a scene.

2. Related Art

Range imaging systems are used in a variety of applications to determine the three-dimensional (3-D) characteristics of a scene (a scene is an environment of interest). By way of example, these applications include 3-D scene reconstruction, 3-D object recognition, robot navigation, terrain mapping and object tracking. An important component of a range imaging system is a range camera. A range camera is a device that is used to measure a 3-D structure of a scene by providing range (or depth) information as measured from a plane on the camera. Thus, while a black and white camera provides a grayscale intensity of each pixel and a color camera provides a color of each pixel, a range camera provides a range (or distance to the 3-D scene) of each pixel. Range cameras use a variety

of techniques to measure range including lasers, projected light patterns and stereo vision.

For some applications (such as tracking persons within a scene) the range imaging system may include more than one range camera because a single range camera may not have a sufficiently large field of view to monitor the entire scene. In order for multiple range cameras to work together, however, the cameras must be calibrated to determine a position and an orientation of each camera relative to one of the cameras (known as a relative pose). This calibration of multiple cameras enables the ranging system to convert 3-D measurements obtained from each camera into a common coordinate frame. For example, a path of a person in a scene may be measured by each camera in its local coordinate frame and converted to a common coordinate frame (such as a room-based coordinate system).

Several types of manual calibration techniques are used to calibrate the range cameras. One type of calibration technique uses a three-dimensional calibration chart to determine the relative position of each camera. This technique, however, is difficult to use and time-consuming because it requires that the calibration chart be positioned correctly within a scene.

Another type of calibration technique requires a user to monitor a scene and determine a plurality of reference points in the scene until the relative position of each camera can be determined. For example, a user references a number of common points in a scene (within each camera's field of view) and, if enough of these common points are found, the relative pose of the cameras may be determined. One disadvantage of this technique, however, is that it is difficult to implement in a consumer-based product because it is unlikely the consumer would want to perform such a complicated and time-consuming calibration process. Moreover, with both

types of calibration techniques, if the consumer performed the calibration process improperly any results obtained from the range imaging system would be erroneous.

Accordingly, there exists a need for a range camera calibration  
5 method and system that is accurate and simple to use. Whatever the merits of the above-mentioned systems and methods, they do not achieve the benefits of the present invention.

### SUMMARY OF THE INVENTION

10 To overcome the limitations in the prior art as described above and other limitations that will become apparent upon reading and understanding the present specification, the present invention includes a method and system for determining a relative position and orientation of a plurality of range cameras using spatial movement. In particular, a path of  
15 an object is measured by each range camera in the camera's local coordinate frame. Thus, the path of the object is observed by each camera but, because each camera has a different view of the object's path, the object path is reported by each camera in different local coordinate frames.

20 The present invention determines the relative location of each range camera by converting the object path as measured in each of the local coordinate frames to a common coordinate frame. The common coordinate frame may be, for example, with respect to one of the cameras or with respect to the scene (such as a room-based coordinate system).

25 In general, the novel method of the present invention includes measuring a path of an object in a scene as observed by each camera, performing matching of points of the path and obtaining transformation parameters (such as an offset distance ( $\Delta x$ ,  $\Delta y$ ) and a rotation angle ( $\theta$ )), preferably by solving a system of transformation equations. These

transformation parameters are used to determine the relative position of each camera. Moreover, the present invention includes other novel features such a data synchronization feature that uses a time shift between cameras to obtain the transformation parameters. In addition, the present

5 invention includes a unique process that improves the robustness and accuracy of solving the system of transformation equations by using a process that is less sensitive to outlying points. For example, in a preferred implementation the present invention includes using a least median of squares technique to reduce the sensitivity of the solution to

10 points extremely removed from the correct solution. The present invention also includes an interpolation process that interpolates between sampled points if there is no data at a particular instant in time. Further, the present invention includes a system for determining a relative position and orientation of range cameras using spatial movement that incorporates the

15 method of the present invention.

Other aspects and advantages of the present invention as well as a more complete understanding thereof will become apparent from the following detailed description, taken in conjunction with the accompanying drawings, illustrating by way of example the principles of the invention.

20 Moreover, it is intended that the scope of the invention be limited by the claims and not by the preceding summary or the following detailed description.

#### BRIEF DESCRIPTION OF THE DRAWINGS

25 The present invention can be further understood by reference to the following description and attached drawings that illustrate the preferred embodiments. Other features and advantages will be apparent from the following detailed description of the invention, taken in conjunction with the

accompanying drawings, which illustrate, by way of example, the principles of the present invention.

Referring now to the drawings in which like reference numbers represent corresponding parts throughout:

5 FIG. 1 is a block diagram illustrating an apparatus for carrying out the present invention.

FIG. 2 is an overall block diagram of a range imaging system incorporating the present invention.

10 FIG. 3 is a general block diagram of the object tracker of the present invention.

FIG. 4 is a block diagram illustrating the calibration module of the object tracker shown in FIG. 3.

FIG. 5 is a block diagram illustrating the transformation processor the calibration module shown in FIG. 4.

15 FIG. 6 is a general flow diagram of the operation of the calibration module of the present invention.

FIG. 7A-7C are general block diagrams illustrating exemplary operations of the calibration module shown in FIG. 4.

20 FIG. 8 is a detailed flow diagram illustrating a preferred embodiment of the present invention.

FIGS. 9A-9D illustrate an example of how the present invention can perform data matching.

FIG. 10 illustrates an example of how the present invention can determine an accurate time offset value.

25

#### DETAILED DESCRIPTION OF THE INVENTION

In the following description of the invention, reference is made to the accompanying drawings, which form a part thereof, and in which is shown by way of illustration a specific example whereby the invention may be

practiced. It is to be understood that other embodiments may be utilized and structural changes may be made without departing from the scope of the present invention.

5    I.    Exemplary Operating Environment

FIG. 1 and the following discussion are intended to provide a brief, general description of a suitable computing environment in which the invention may be implemented. Although not required, the invention will be described in the general context of computer-executable instructions, such as program modules, being executed by a computer. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. Moreover, those skilled in the art will appreciate that the invention may be practiced with a variety of computer system configurations, including personal computers, server computers, hand-held devices, multiprocessor systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, and the like. The invention may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located on both local and remote computer storage media including memory storage devices.

With reference to FIG. 1, an exemplary system for implementing the invention includes a general-purpose computing device in the form of a conventional personal computer 100, including a processing unit 102, a system memory 104, and a system bus 106 that couples various system components including the system memory 104 to the processing unit 102. The system bus 106 may be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local

bus using any of a variety of bus architectures. The system memory includes read only memory (ROM) 110 and random access memory (RAM) 112. A basic input/output system (BIOS) 114, containing the basic routines that help to transfer information between elements within the

5 personal computer 100, such as during start-up, is stored in ROM 110.

The personal computer 100 further includes a hard disk drive 116 for reading from and writing to a hard disk, not shown, a magnetic disk drive 118 for reading from or writing to a removable magnetic disk 120, and an optical disk drive 122 for reading from or writing to a removable optical disk

10 124 such as a CD-ROM or other optical media. The hard disk drive 116, magnetic disk drive 128 and optical disk drive 122 are connected to the system bus 106 by a hard disk drive interface 126, a magnetic disk drive interface 128 and an optical disk drive interface 130, respectively. The drives and their associated computer-readable media provide nonvolatile

15 storage of computer readable instructions, data structures, program modules and other data for the personal computer 100.

Although the exemplary environment described herein employs a hard disk, a removable magnetic disk 120 and a removable optical disk 124, it should be appreciated by those skilled in the art that other types of computer readable media that can store data that is accessible by a computer, such as magnetic cassettes, flash memory cards, digital video disks, Bernoulli cartridges, random access memories (RAMs), read-only memories (ROMs), and the like, may also be used in the exemplary operating environment.

20 25 A number of program modules may be stored on the hard disk, magnetic disk 120, optical disk 124, ROM 110 or RAM 112, including an operating system 132, one or more application programs 134, other program modules 136 and program data 138. A user (not shown) may enter commands and information into the personal computer 100 through

input devices such as a keyboard 140 and a pointing device 142. In addition, a camera 143 (or other types of imaging devices) may be connected to the personal computer 100 as well as other input devices (not shown) including, for example, a microphone, joystick, game pad, 5 satellite dish, scanner, or the like. These other input devices are often connected to the processing unit 102 through a serial port interface 144 that is coupled to the system bus 106, but may be connected by other interfaces, such as a parallel port, a game port or a universal serial bus (USB). A monitor 146 or other type of display device is also connected to 10 the system bus 106 via an interface, such as a video adapter 148. In addition to the monitor 146, personal computers typically include other peripheral output devices (not shown), such as speakers and printers.

The personal computer 100 may operate in a networked environment using logical connections to one or more remote computers, 15 such as a remote computer 150. The remote computer 150 may be another personal computer, a server, a router, a network PC, a peer device or other common network node, and typically includes many or all of the elements described above relative to the personal computer 100, although only a memory storage device 152 has been illustrated in FIG. 1.

20 The logical connections depicted in FIG. 1 include a local area network (LAN) 154 and a wide area network (WAN) 156. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets and the Internet.

When used in a LAN networking environment, the personal 25 computer 100 is connected to the local network 154 through a network interface or adapter 158. When used in a WAN networking environment, the personal computer 100 typically includes a modem 160 or other means for establishing communications over the wide area network 156, such as the Internet. The modem 160, which may be internal or external, is

connected to the system bus 106 via the serial port interface 144. In a networked environment, program modules depicted relative to the personal computer 100, or portions thereof, may be stored in the remote memory storage device 152. It will be appreciated that the network connections shown are exemplary and other means of establishing a communications link between the computers may be used.

## II. Introduction

The method and system of the present invention include measuring the relative position and orientation of at least two range cameras. Range cameras, which are used to measure the 3-D structure of a scene, give the range (or depth) of each pixel. In order for two or more range cameras to work properly together, the system (such as a range imaging system) using the range cameras must be able to determine a relative position and orientation of each camera. This process of determining a relative pose of each camera (also known as calibration) enables the system to convert 3-D measurements from each camera into a common coordinate frame. Data from each camera is in the camera's local coordinate frame, and calibration of each camera makes the 3-D measurements from different cameras (in different local coordinate frames) consistent with each other.

The present invention measures a relative pose between a plurality of range cameras by measuring a relative pose between two cameras at a time. One camera is designated as a base camera and relative poses of the remainder of the cameras can be measured relative to the base camera. In general, the present invention calibrates range cameras based on a path of an object around a scene. The object path is determined in a ground plane (such as a floor of a room) as a function of time as measured by a range camera. The present invention determines the transformation parameters that take a point on the object path measured by a non-base

camera and convert it to a point as it would be seen from the base camera. In addition, the present invention includes synchronizing data obtained from each camera, interpolating between sampled data points and using a robust error minimization technique to determine the transformation 5 parameters.

### III. General Overview

As shown in FIGS. 2-10 for the purposes of illustration, the invention is embodied in a method and a system for measuring a relative position 10 and orientation of range cameras using a movement of an object within a scene. FIG. 2 is an overall block diagram of a range imaging system incorporating the present invention. The range imaging system illustrated is only one example of several systems that could incorporate the relative range camera calibration method and system of the present invention. In 15 general, the range imaging system 200 includes a first camera 208 and a second camera 216. Each of the cameras 208, 216 may use any of the various techniques available to measure range, such as, for example, lasers, projected light patterns and stereo vision. Both of the cameras 208, 216 are directed toward a scene 224 and are capable of measuring a 3-D 20 structure of the scene 224.

The range imaging system also includes a first data module 232 that samples raw position data from the first camera 208 and a second data module 236 that samples raw position data from the second camera 216. These data modules 232, 236 may be, for example, computers or 25 microprocessors. The first camera 208 supplies position data about the scene 224 in a first local coordinate frame and the second camera 216 supplies position data about the scene 224 in a second local coordinate frame. These two local coordinate frames generally are not the same, and

calibration of the two cameras 208, 216 is necessary to express the position data from each camera in a common coordinate frame.

The sampled data from each camera is sent to an object tracker 240, which inputs the sampled data, calibrates the cameras 208, 216 and 5 performs a coordinate transformation of the data. Further, an output module 248 is included in the range imaging system 200 that outputs scene data in a common coordinate system (such as a room-based coordinate system). In this example, the scene 224 includes a room 256 containing a first sofa 264 on one side of the room 256 and a second sofa 10 272 opposite the first sofa 264. In addition, a chair 280 is situated between to sofas 264, 272.

In this range imaging system, calibration of the range cameras 208, 216 generally is performed by having a person 288 (denoted by an "X") move in a path 296 around the room 256. This path 296 is observed by 15 the cameras 208, 216 in their respective local coordinate frames and the raw position data (such as (x,y) coordinates) of the path 296 is sampled by the data modules 232, 236. The data modules 232, 236 sample raw position data from each camera that includes the object path 296 described in a first local coordinate frame (as observed by the first camera 208) and the object path 296 described in a second local coordinate frame 20 (as observed by the second camera 216).

The object tracker 240 receives the sampled data from the data modules 232, 236 and, using the present invention, calibrates cameras 208, 216 by determining the relative position and orientation of each 25 camera. Once the cameras 208, 216 are calibrated any data from the cameras 208, 216 is converted into a common coordinate frame. This means, for example, a path of an object around the room 256 is expressed by the object tracker 240 in a common coordinate frame. The object tracker 240 sends data in a common coordinate frame to the output

module 248, for output from the range imaging system 200. Further, the range imaging system 200 may transmit the data to a post-processing module 298 that may include, for example, a three-dimensional (3-D) scene reconstruction system, a 3-D object recognition system or a 3-D 5 tracking system (which may be part of a vision-based computer interface system).

#### IV. Component Overview

FIG. 3 is a general block diagram of the object tracker 300 (the object tracker 240 in FIG. 2 is one example of the object tracker 300) of the present invention. In general, position data from cameras (box 310) in local coordinate frames is received by the object tracker 300, processed and data is sent as output in a common coordinate frame (box 320). The object tracker 300 includes a calibration module 330, which determines transformation parameters that will transform position data in local coordinate frames into a common coordinate frame, and a coordinate processor 340, which uses the transformation parameters computed by the calibration module 330 to transform data observed by the cameras into a desired common coordinate frame.

FIG. 4 is a block diagram illustrating the calibration module 330 of the object tracker 300 shown in FIG. 3. The calibration module determines transformation parameters that are used to convert data in a local coordinate frame of each camera into a common coordinate frame. In general, data from each camera is received as input (box 410) and a data synchronizer 420 is used to synchronize the data received from multiple cameras. A coordinate selector 430 determines the desired coordinate frame of the transformation. For example, a first camera may be selected as the base camera and data from the other cameras are expressed in the coordinate frame of the base camera. A transformation processor 440

computes transformation parameters that convert data from a local coordinate frame to be expressed in the base coordinate frame. These transformation parameters are sent as output (box 450) of the calibration module 330.

5 FIG. 5 is a block diagram illustrating the transformation processor 440 the calibration module shown in FIG. 4. The transformation processor 440 includes an interpolation module 510, for interpolating between data points, a data matching processor 520, for matching up data points from different cameras at a certain time, and an error minimization processor 10 530, for determining the data points that yield the most accurate transformation parameters. The transformation processor 440 inputs synchronized data from the data synchronizer 420. A time is then selected by the interpolation module 510 along with position data corresponding to that time. If there was no data point sampled by the data modules at the 15 selected time then the interpolation module 510 interpolates a data point, as described further below.

The data points at the selected time are received by the data matching processor 520. In addition, the data matching processor 520 receives a desired coordinate frame as determined by the coordinate selector 430. The desired coordinate frame may be, for example, chosen 20 by the user or selected at random. Any data from the cameras is expressed in the selected coordinate frame (also called the base coordinate frame). The data matching processor 520 matches data points at the selected time and computes transformation parameters using the 25 data points. The error minimization processor 530 determines which data points give the most accurate transformation parameters.

V. Details of the Components and Operation

FIG. 6 is a general flow diagram of the operation of the calibration module 330 of the present invention. Generally, data observed by cameras is received as input (box 610). Depending on the application there may be two or more cameras, with each camera positioned to observe data within a scene. Even if there are more than two cameras, however, the calibration method of the present invention only needs to measure the relative position and orientation between two cameras at a time. This is because one camera is designated as the base camera and the position and orientation of the remainder of the cameras are measured from that base camera. The raw data from each camera is sampled and sent to the calibration module 330 and the sampled data from each camera in its respective local coordinate frame.

One of the cameras is selected as the base camera and the coordinate frame of chosen camera becomes the base coordinate frame (box 620). Transformation parameters are computed (box 630) from the sampled data received by the calibration module 330. These transformation parameters are then used to express data received from each camera in the base coordinate frame. Once this calibration process is performed, any data observed by a non-base camera can be expressed in the base coordinate frame as if the data had been observed by the base camera.

FIG. 7A-7C are general block diagrams illustrating the operations of the transformation parameters computed by the calibration module 330 shown in FIG. 4. In FIG. 7A, a first camera 700 and a second camera 705 observe an object path in a scene. A first object path 710 is observed by the first camera 700 in a first local coordinate frame and a second object path 715 is observed by the second camera in a second local coordinate

frame. It should be noted that two cameras 700, 705 observe the same path but in different local coordinate frames.

FIG. 7B illustrates a set of transformation parameters computed by the calibration module 330 applied to the data of the second camera 705.

5 In particular, local coordinate frame of the first camera 700 has been selected as the base coordinate frame and, in accordance with the present invention, one purpose of the calibration module 330 is to compute transformation parameters that cause the second object path 715 to overlap with the first object path 710 as closely as possible. The

10 transformation parameters include a change in the "x" coordinate ( $\Delta x$ ) 730, a change in the "y" coordinate ( $\Delta y$ ) 735 and an angle of rotation ( $\theta$ ) 740. As shown in FIG. 7B, when the transformation parameters ( $\Delta x$ ,  $\Delta y$ ,  $\theta$ ) are applied to the second object path 715, the first object path 710 and second object path 715 nearly overlap. The lack of exact overlap is due to a slight

15 amount of error in the calculation of the transformation parameters.

FIG. 7C illustrates another set of transformation parameters computed by the calibration module 330 applied to the data of the second camera 705. In FIG. 7C the transformation parameters  $\Delta x^*$  750,  $\Delta y^*$  755 and  $\theta^*$  760 are used to achieve an exact overlap of the first object path 710 and the second object path 715 into a single object path 770. The exact overlap represents minimum error in the transformation parameters ( $\Delta x^*$ ,  $\Delta y^*$ ,  $\theta^*$ ) and means that these transformation parameters can be used to express data from the second camera 705 in the base coordinate frame.

25 FIG. 8 is a detailed flow diagram illustrating a preferred embodiment of the present invention. In this preferred embodiment, the present invention designates one of a plurality of cameras as a base camera and measures the relative pose of the remainder of the cameras with respect to the base camera. Initially, one camera is selected as a base camera and

that camera's local coordinate frame becomes the base coordinate frame (box 800). Moreover, data from each camera is received as input (box 805).

Before this data can be used to compute transformation parameters, 5 however, at least two problems must be overcome. The first problem occurs if the clocks on the computers used to sample the data are unsynchronized by a constant time offset so that equivalent time readings on the computers do not correspond to the same actual time. The present invention corrects this problem by adding a time offset to the data. In 10 particular, a time offset value is chosen (box 810) and applied to the camera data (box 815) in order to synchronize the data. The second problem occurs if the data from the cameras is not sampled at the same time leaving, for example, a data point at time  $t$  from a first camera without a corresponding data point from a second camera.

15 The present invention corrects this problem by performing a linear interpolation (box 820) between two data points sampled before and after time  $t$ . This linear interpolation approximates where a data point would have been seen at time  $t$ . Next, data matching is performed to provide enough data points to compute the corresponding transformation 20 parameters. Data matching matches data from different cameras at certain absolute times and uses these data points to compute transformation parameters.

FIGS. 9A-9D illustrate an example of how the present invention can perform data matching. In particular, data from a first camera (camera 1) 25 in a first local coordinate frame  $(x_1, y_1)$  and data from a second camera (camera 2) in a second coordinate frame  $(x_2, y_2)$  are graphed as a function of time. FIG. 9A illustrates a graph of  $x_1$  versus time, FIG. 9B illustrates a graph of  $y_1$  versus time, FIG. 9C illustrates  $x_2$  versus time and FIG. 9D illustrates  $y_2$  versus time. A time  $T^*$  is selected such that there is data

available at  $T^*$  from, for example, camera 1. At time  $T^*$ , the  $x_1$  and  $y_1$  coordinates from camera 1 (assuming that camera 1 was selected) will be perfectly synchronized, but there may be no data (i.e.  $(x_2, y_2)$ ) available from camera 2. In this situation, data from camera 1 at time  $T^*$  is matched

5 such that a first point 910 on the  $x_1$  versus time graph (or a second point 920 on the  $y_1$  versus time graph) are matched with a third point 930 on the  $x_2$  versus time graph and a fourth point 940 on the  $y_2$  versus time graph.

Note that in FIGS. 9C and 9D there are no sampled data points from camera 2 at time  $T^*$ . The present invention performs an interpolation and

10 chooses sampled data points 950, 955 prior to  $T^*$  and sampled data points 960, 965 after  $T^*$ . These sampled points 950, 955, 960, 965 are used to interpolate values of  $x_2$  and  $y_2$  at time  $T^*$  to obtain the third and fourth data points 930, 940. Once the data has been matched at a certain time, the invention determines whether more data points are needed (box 830). If

15 so, then a different time is chosen and data matching is performed (box 825) at that time. Otherwise, if there are enough data points, an error minimization technique is used to find the data points that give the transformation parameters with the least error (box 835).

For example, the present invention may determine minimum error by

20 using a least squares technique that is discussed by S. Ma and Z. Zhang in "Computer Vision" (Chinese Academy of Science, 1998), the entire contents of which are hereby incorporated by reference. In a preferred embodiment, however, the present invention uses a least median of squares technique to determine minimum error. The least median of

25 squares technique is more robust and less affected by data points that lie well away from the majority of data points. The least median of squares technique is discussed in detail by P.J. Rousseeuw and A.M. Leroy in "Robust Regression and Outlier Detection" (New York: John Wiley and

Sons, 1987), the entire contents of which are hereby incorporated by reference.

When the transformation parameters with the least amount of error have been determined, they are stored along with the time offset value 5 used to synchronize the data (box 840). Next, a determination is made whether more time offset points are needed (box 845). If more are needed, then another time offset value is selected (box 850) and the process begins again at box 815. Otherwise, an error minimization technique is used to find the time offset value with the least amount of 10 error (box 855). As before, the least median of squares technique is a preferred technique to determine the minimum error.

FIG. 10 illustrates an example of how the present invention can determine an accurate time offset. Specifically, the error for each time offset values is determined and plotted as shown in FIG. 10. A point 1000 15 at which a minimum error occurs a corresponding time offset value is noted. When the time offset value at minimum error is determined, both the time offset value and the corresponding transformation parameters are sent as output (box 860).

It should be noted that in a preferred embodiment the transformation 20 parameters are changes in the x and y coordinates and the rotation angle (such as  $\Delta x$ ,  $\Delta y$ ,  $\theta$ ). In addition, other transformation parameters may be used depending on the type of coordinate systems used (such as, for example, polar coordinate systems).

25 VI. Working Example

The following working example uses a range imaging system to track the movement of a person around a room and is provided for illustrative purposes only. In this working example, the method and system of the present invention are used to calibrate two range cameras in prior to

using the range imaging system. As mentioned above, a variety of techniques (such as lasers and projected light patterns) are available for measuring range. Although in general the present invention is capable of using any ranging technique, in this working example stereo cameras were

5 used. Stereo cameras were chosen because of their fast frame rate and because they are inexpensive and safe. In this working example, the application was tracking people as they move around a room. Further, two range cameras (camera 1 and camera 2) were used and calibrated based on a person's path when the person walked around the room.

10 The calibration process began by determining an (x,y) location of the person on a ground plane (in this working example, the floor of the room) as a function of time as measured by each range camera. This was accomplished using a technique described in co-pending U.S. patent application serial number 09/455,822 entitled "A System and Process for

15 Locating and Tracking a Person or Object in a Scene Using a Series of Range Images" by Barry Brumitt, filed on December 6, 1999, the entire contents of which are hereby incorporated by reference. The present invention then chose a first camera as the base camera and designated the location measured by the base camera as  $(x_1, y_1)$  and a corresponding

20 point from a second (non-base) camera (camera 2) as  $(x_2, y_2)$ . The present invention was used to calibrate the two cameras by computing the transformation parameters of an angle  $\theta$  and an offset  $(\Delta x, \Delta y)$  that made the following equation true:

$$\begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & \Delta x \\ \sin(\theta) & \cos(\theta) & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix}$$

25

Once the values of  $\theta$  and  $(\Delta x, \Delta y)$  were determined, using this equation, any point seen by camera 2,  $(x_2, y_2)$  could be transformed into the coordinates of camera 1  $(x_1, y_1)$ . This means that a point as seen by camera 2 could be expressed in the local coordinate frame of camera 1 as if the point was  
5 actually seen by camera 1.

In order to synchronize the data coming from the first and second cameras, a time offset  $\Delta t$  was used to correct for the fact that a clock on the computer associated with the first camera (clock 1) was not synchronized with a clock on the computer associated with the second  
10 camera (clock 2). Thus, the points from the first camera and the second camera became  $(x_{1i}, y_{1i}, t_{1i})$  and  $(x_{2j}, y_{2j}, t_{2j} + \Delta t)$ , respectively. An initial guess of the time offset  $\Delta t$  was chosen and a point from each camera was sampled. Because the sampled points from each camera did not exactly match up with each other, the data obtained from the second camera was  
15 interpolated as follows.

First, for every point from the first camera taken at time  $t_{1i}$ , two points from the second camera were found that were taken as close as possible on either side of that time (i.e., points  $j^-$  and  $j^+$  were found such that  $t_{2j^-} + \Delta t < t_{1i} < t_{2j^+} + \Delta t$ ). Next, a linear interpolation was performed on the  
20 two points from the second camera,  $(x_{2j^-}, y_{2j^-})$  and  $(x_{2j^+}, y_{2j^+})$ , to approximate where the point would have been had it been seen at time  $t_{1i}$ . If, for any point in the first data set, surrounding points in the second data set could not be found, that point in time was ignored. After ignoring such points and after interpolation, there was a set of corresponding  $(x, y)$  points that were  
25 designated as  $(x_{1k}^*, y_{1k}^*, x_{2k}^*, y_{2k}^*)$ ,  $1 \leq k \leq n$ . The time data in this data set was ignored because it made no difference in the subsequent computations.

Next, point matching was performed and the transformation parameters corresponding to the least squared error was selected. Specifically, in this working example the least median of squares technique was used because it is a robust method. This method was implemented

5 by picking random pairs of corresponding points from the data set  $((x_{1k}^*, y_{1k}^*, x_{2k}^*, y_{2k}^*), 1 \leq k \leq n)$ . A pair of points was the minimum number needed to compute the candidate transformation parameters (i.e.,  $\theta$  and  $(\Delta x, \Delta y)$ ).

The two pairs of randomly chosen points were

$(x_{1a}^*, y_{1a}^*), (x_{1b}^*, y_{1b}^*), (x_{2a}^*, y_{2a}^*), (x_{2b}^*, y_{2b}^*)$ , and the angle  $\theta$  was computed as:

$$\begin{aligned}\Delta x_a &= x_{2a}^* - x_{1a}^* \\ \Delta y_a &= y_{2a}^* - y_{1a}^* \\ \Delta x_b &= x_{2b}^* - x_{1b}^* \\ \Delta y_b &= y_{2b}^* - y_{1b}^* \\ \cos(\theta) &= \frac{\Delta y_a \Delta y_b + \Delta x_a \Delta x_b}{\Delta x_b \Delta x_b + \Delta y_b \Delta y_b} \\ \sin(\theta) &= \frac{\Delta x_b \Delta y_a - \Delta x_a \Delta y_b}{\Delta x_b \Delta x_b + \Delta y_b \Delta y_b} \\ \theta &= \tan^{-1}\left(\frac{\sin(\theta)}{\cos(\theta)}\right)\end{aligned}$$

and the translation  $(\Delta x, \Delta y)$  was:

$$\begin{aligned}\Delta x &= x_{1a}^* - x_{2a}^* \cos(\theta) + y_{2a}^* \sin(\theta) \\ \Delta y &= y_{1a}^* - x_{2a}^* \sin(\theta) - y_{2a}^* \cos(\theta)\end{aligned}$$

This  $\theta$  and  $(\Delta x, \Delta y)$  served as a trial solution for the calibration

15 problem based on the two randomly chosen pair of points. The solution was evaluated by computing a list of the squared errors between corresponding points:

$$e_k = (x_{1k}^* - x_{2k}^* \cos(\theta) + y_{2k}^* \sin(\theta) - \Delta x)^2 + (y_{1k}^* - x_{2k}^* \sin(\theta) - y_{2k}^* \cos(\theta) - \Delta y)^2$$

The quality of the solution was the median value of this list of squared errors. In this working example, our implementation, 100 random pairs of

corresponding points were chosen and the transformation parameters  $\theta$  and  $(\Delta x, \Delta y)$  that corresponded to the least median of squares were used.

The least median of square technique was used as above to compute the best  $\theta$  and  $(\Delta x, \Delta y)$  for a whole series of values of a time offset value ( $\Delta t$ ). Whichever  $\Delta t$  gave the minimum least median of squares was chosen as the best one, and the corresponding  $\theta$  and  $(\Delta x, \Delta y)$  were used for the final solution.

As an alternative to the least median of square technique described above, a least square solution could have been used to determine a minimum error. The least squares solution to the calibration problem computes the transformation parameters  $\theta$  and  $(\Delta x, \Delta y)$  that minimize the sum of the Euclidean distances between corresponding points in  $(x_{1k}^*, y_{1k}^*, x_{2k}^*, y_{2k}^*)$ ,  $1 \leq k \leq n$ . The angle,  $\theta$ , is given by

$$\tan(\theta) = \frac{\sum_{k=1}^n ((x_{2k}^* - \bar{x}_2^*)(y_{1k}^* - \bar{y}_1^*) - (y_{2k}^* - \bar{y}_2^*)(x_{1k}^* - \bar{x}_1^*))}{\sum_{k=1}^n ((x_{1k}^* - \bar{x}_1^*)(x_{2k}^* - \bar{x}_2^*) + (y_{1k}^* - \bar{y}_1^*)(y_{2k}^* - \bar{y}_2^*))}$$

The above equation depends on the following equation, which computes the centroids of the points from each camera

$$(\bar{x}_1^*, \bar{y}_1^*) = \frac{1}{n} \left( \sum_{k=1}^n x_{1k}^*, \sum_{k=1}^n y_{1k}^* \right)$$

$$(\bar{x}_2^*, \bar{y}_2^*) = \frac{1}{n} \left( \sum_{k=1}^n x_{2k}^*, \sum_{k=1}^n y_{2k}^* \right)$$

The translation  $(\Delta x, \Delta y)$  is then given by

$$\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} \bar{x}_1^* \\ \bar{y}_1^* \end{pmatrix} - \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{pmatrix} \bar{x}_2^* \\ \bar{y}_2^* \end{pmatrix}$$

The  $\theta$  and  $(\Delta x, \Delta y)$  computed are the solution to the calibration problem.

The quality (or amount of error) of the solution is given by the average squared distance between corresponding points:

$$e_2 = \frac{1}{n} \sum_{k=1}^n \left[ (x_{1k} - x_{2k} \cos(\theta) + y_{2k} \sin(\theta) - \Delta x)^2 + (y_{1k} - x_{2k} \sin(\theta) - y_{2k} \cos(\theta) - \Delta y)^2 \right]$$

For a series of values of the time offset,  $\Delta t$ , the transformation parameters  $\theta$ ,  $(\Delta x, \Delta y)$  and  $e_2$  are computed. The average squared distance between corresponding points,  $e_2$ , will be a minimum for some value of  $\Delta t$ . We take

5 the corresponding values of  $\theta$  and  $(\Delta x, \Delta y)$  at the minimum value of  $\Delta t$  as the solution to the calibration problem.

This least squares solution works well in spite of small errors in tracking the position of the person in the room. However, there can be outlier points due to gross errors in the process that tracks the person.

10 These outlier points are  $(x, y)$  locations that deviate greatly from the actual location of the person. In this case, the least squares solution will be drawn away from the right answer, and a technique that is robust to such errors should be used, such as the least median of square technique described above.

15 The foregoing description of the preferred embodiments of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching. It is intended that the scope of the invention  
20 be limited not by this detailed description of the invention, but rather by the claims appended hereto.